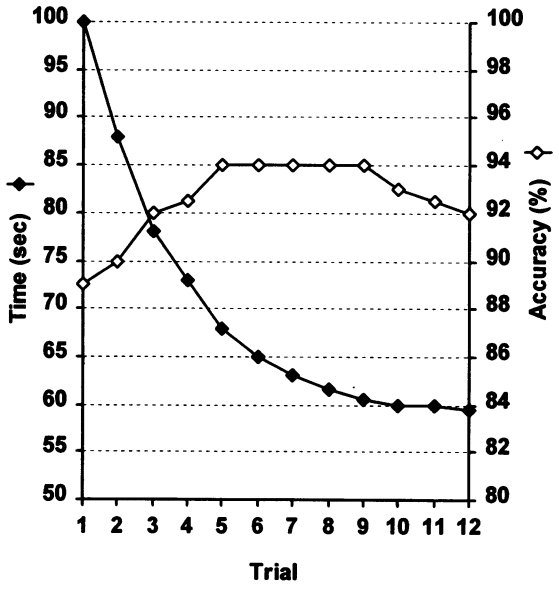
Learning Curves for Speech Recognition in Family Practice

Robert R. Hausam, M.D.[1], Howard Wing, M.D.[2]
Department of Family and Preventive Medicine[1,2]
Office of Information Resources[1]
University of Utah School of Medicine
Salt Lake City, Utah

The ability to automatically recognize and transcribe speech for clinical documentation has been desired and promised for many years. This capability may be especially useful in family practice, due to the relatively high patient volume and the diversity of problems that may be encountered. Unfortunately, most successful implementations of speech recognition have been in other fields, such as radiology, emergency medicine, anesthesiology, pathology, and for some specific procedures such as endoscopy. These areas tend to be characterized by fairly well-defined vocabularies and highly structured and predictable documentation formats. This is not the case in family practice or other primary care fields.

This paper reports our preliminary findings in applying currently available speech recognition technology to the problem of clinical documentation in family practice. For the study, we used the IBM VoiceType™ speech recognition system. This system does not currently have a vocabulary specifically tailored for family practice or other primary care fields. Instead we used their Emergency Medicine supplemental vocabulary. The VoiceType™ system requires discrete speech input, which does not allow the provider to dictate in a normal continuous manner. The process of adaptation of the provider to this mode of operation requires a learning process, which can be characterized by a learning curve.

We utilized a single moderately complex chart note in SOAP format of a typical upper respiratory infection. The same provider repeatedly dictated the note a total of 12 times. For each case we measured the time required for dictation and recorded the transcription accuracy. From this data we generated the following learning and accuracy curves.



This data shows that provider dictation time decreases in a nearly exponential fashion in the initial learning period. In addition, it appears that there may be an optimum dictation rate which minimizes speech recognition error. We intend to extend this work in the future using multiple providers and multiple chart notes for a variety of medical problems in order to verify the reproducibility of these findings.